

SCIENTIFIC REPORTS



OPEN

Causal Inference in the Perception of Verticality

Ksander N. de Winkel , Mikhail Katliar, Daniel Diers & Heinrich H. Bühlhoff 

The perceptual upright is thought to be constructed by the central nervous system (CNS) as a vector sum; by combining estimates on the upright provided by the visual system and the body's inertial sensors with prior knowledge that upright is usually above the head. Recent findings furthermore show that the weighting of the respective sensory signals is proportional to their reliability, consistent with a Bayesian interpretation of a vector sum (Forced Fusion, FF). However, violations of FF have also been reported, suggesting that the CNS may rely on a single sensory system (Cue Capture, CC), or choose to process sensory signals based on inferred signal causality (Causal Inference, CI). We developed a novel alternative-reality system to manipulate visual and physical tilt independently. We tasked participants ($n = 36$) to indicate the perceived upright for various (in-)congruent combinations of visual-inertial stimuli, and compared models based on their agreement with the data. The results favor the CI model over FF, although this effect became unambiguous only for large discrepancies ($\pm 60^\circ$). We conclude that the notion of a vector sum does not provide a comprehensive explanation of the perception of the upright, and that CI offers a better alternative.

Whenever we specify objects' relative locations using terms as 'above' or 'below', or when we move throughout the world while trying not to topple over, we make use of the fact that we have a perception of upright. Multiple sensory systems throughout the body provide the nervous system with information that can potentially be used to construct a subjective vertical: visually, we are able to determine our orientation from polarity information in the optic array¹; our vestibular system is stimulated by accelerations, and therefore provides us with information on the direction and magnitude of the gravitational vector²; we receive information about our orientation relative to gravity from pressure cues on the body^{3,4} and the distribution of fluids in the body^{5,6}; and there is evidence for specialized graviceptors located in the trunk⁷.

Mittelstaedt⁸ proposed that the Central Nervous System (CNS) constructs perceptions of verticality by combining the sensory information from the visual system and the body's collective inertial sensors with the prior knowledge that 'up' is usually aligned with the long-body axis (the idiotropic vector), and that the process could be described as a vector sum, where the length of the vectors represents the relative influence of each component. In subsequent work, this concept has been interpreted as a reflection of statistically optimal behavior by the Central Nervous System (CNS): according to Bayes' rule, if sensory estimates of the upright are normally distributed random variables and the prior is either normally distributed or uninformative, the estimate that is most likely the true upright can be calculated as a weighted average of the sensory estimates, where the weights are proportional to the inverse of estimates' variances^{9–13}.

Several studies report that people's perception of the upright indeed reflects such Bayesian integration^{14–18}. Different measures were used among studies: participants were instructed to either indicate the Subjective Visual Vertical (SVV) by aligning an object in the visual display with the perceived upright^{14,16–18}; the perceptual upright was inferred from participants' interpretations of the ambiguous symbol 'p', which is defined by its orientation relative to the perceived upright (equivalently, 'd'; Oriented CHAracter Recognition Test, or OCHART^{14,19}); and estimates of upright have been derived from a discrimination task, where participants discriminated between roll stimuli on the basis of Subjective Body Tilt¹⁷ (SBT). Although biases and intersensory weightings appear to differ between tasks²⁰, the ratios of the weights attributed to the constituent cues coincided well with calculations of sensory variance for the former measures, thereby providing supporting evidence for the Bayesian interpretation of the vector-sum model; and for the latter measure, supporting evidence was obtained by comparisons of model fit indices.

Department of Human Perception, Cognition, and Action, Max Planck Institute for Biological Cybernetics, Max-Planck-Ring 8, 72076, Tübingen, Germany. Correspondence and requests for materials should be addressed to K.N.W. (email: ksander.dewinkel@tuebingen.mpg.de)

However, some findings are inconsistent with such modeling. First, the weightings reported differ between experiments and appear to depend on the specific measure^{14,18}. From the perspective of a Bayesian vector sum model, this means that the sensory variances differ between tasks and experiments. Even though it is not unlikely that specific conditions of an experiment affect the sensory estimates, it is not clear why the variance of sensory estimates of the upright would vary depending on the task. Second, De Winkel *et al.*²¹ performed a study where participants were asked to report the SVV during exposure to different levels of gravity during parabolic flight. Here, it was found that participants discarded the visual cue entirely, and relied on either inertial or idiosyncratic information in a dichotomous fashion, where the probability of relying on the inertial cue was proportional to the strength of the gravitational pull.

The observed differences in the role specific cues fulfill in different tasks and the aforementioned variability in sensory weightings imply that Bayesian models based on the notion of a vector sum cannot offer a comprehensive explanation of the perception of the upright. Consistent with recent reports on audiovisual interactions in spatial localization tasks^{22–25} and visual-inertial heading estimation^{26,27}, it is possible that the role attributed to different cues reflects the inferred causality of the signals. Models that can account for different behaviors depending on inferred causality are Causal Inference (CI) models^{22,23}. Put simply, these models state that the CNS constructs intermediate estimates of environmental properties consistent with different interpretations of their causes (i.e., a common cause or separate causes) in tandem, and combines these into final estimates, taking into account a-priori beliefs on the probability of alternative causal structures.

We hypothesized that CI models provide a better explanation of the perceptual upright than the vector-sum approach. In an experiment, we independently manipulated participants' physical and visual orientation with respect to the true vertical, and tasked them to provide estimates of what they thought was the true, physical, upright. We developed statistical versions of prominent models of multisensory perception and compared their ability to explain the participants' responses.

Methods

Overview. To test the hypothesis outlined above experimentally, we placed 36 participants on a motion platform that was capable of physically rotating them around their naso-occipital axis. While seated on this platform, participants wore a head-mounted display (HMD) setup that allowed us to manipulate visual orientation independently from the true gravitational vertical. Participants indicated what they thought was upright for various (in)congruent visual-inertial orientation stimuli. We assessed how well a number of alternative models of spatial orientation could account for participants' responses by fitting each model to the data and comparing model fit indices.

Ethics statement. The experiment was carried out in accordance with the declaration of Helsinki. All participants provided written informed consent prior to participation. The experimental protocol was approved by the ethical committee of the medical faculty of the Eberhard-Karls University in Tübingen, Germany, reference number 352/2017BO2.

Data availability statement. All experimental data are available as supplementary material Data S1.

Participants. A total of 40 participants were recruited for the experiment. Two participants were not able to complete the experiment due to motion sickness; data from two other participants had to be excluded due to an inability to perform the task. Of the remaining 36 participants, 20 were male. The mean age was 29.5 years, with a standard deviation of 7.3. Participants 11, 14, and 16 in the first iteration of the experiment also participated as participants 34, 36, and 30 in the third iteration. Participants were compensated for their time at a rate of €8 an hour.

Setup. Visual stimuli were generated and presented using a custom-made alternative-reality setup. This setup allowed us to show participants their immediate visual environment at any desired roll-tilt angle stereoscopically and in real-time, such that the visual stimulus was equivalent to visual tilt experienced due to head tilt. The system was designed with the aim to maximize the ecological validity of the visual stimuli as a cue to orientation; to avoid the possibility that they would be discarded simply because they are unlikely to reflect one's spatial orientation²¹. For instance, when reading a large book while lying on one's side, the orientation of the book does not reflect one's own orientation. Through this system, participants viewed the entrance and control area of the simulator hall. The visible area of the room was approximately 5 m wide by 4 m high and 5 m deep, and included a green crane, stairs, the platform control systems, and the experimenter. The elevated platform is 2 m high, and the nearest edge was 2 m in front of the participant. The visual stimulus could be used to estimate the upright by means of polarity cues, support relationships between objects, and motion of the experimenter (e.g.^{2,28}). A (monoscopic) screenshot of the participants' view is presented in the top panel of Fig. 1.

The alternative-reality system consisted of an OVRVision Pro stereo camera (Wizapply, Osaka, Japan) mounted via a Dynamixel AX12-A servo motor (Robotis, Lake Forest, California, United States) to a Vive HMD (HTC, New Taipei City, Taiwan). Mounting hardware was designed and 3D-printed in-house. The HMD displayed the images of the left and right camera in the respective screens at a rate of 45 frames per second. The screens each have a resolution of 1080 × 1200 px and a field of view of 100 × 110°, corresponding to approximately 11 px per degree. The servo allowed us to manipulate the orientation of the camera with an accuracy of 0.29°. Note that, being equivalent to head-tilt, a clockwise tilt of the camera results in the perception of a counter-clockwise rotation of the visual environment. The software to control the servo and transmit the camera images was also developed in-house. The device is shown in Fig. 1, bottom left panel.

Inertial orientation stimuli were presented using an eMotion 1500 hexapod motion system (Bosch Rexroth AG, Lohr am Main, Germany). For different physical orientation stimuli, the platform was moved in such a way

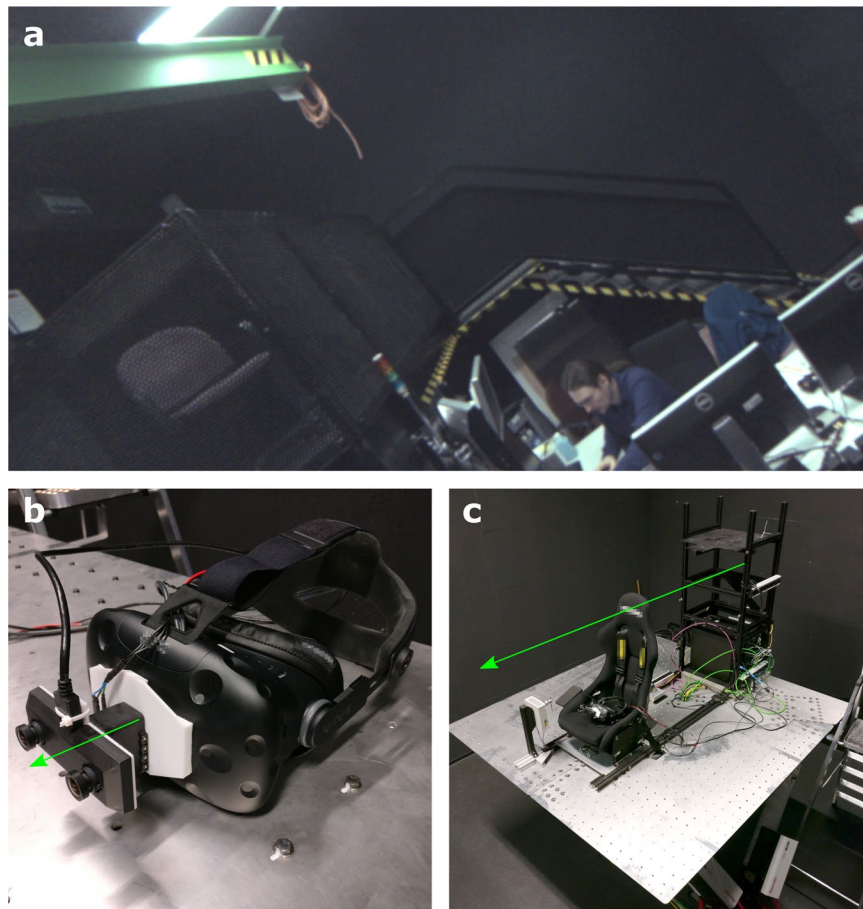


Figure 1. (a) (Monoscopic) screenshot of a participant's view through the alternative-reality system, showing the entrance and control area of the simulator hall. (b) photograph of the alternative-reality system. (c) view of the experimental setup, showing the motion platform and seat. The green arrows indicate axes of rotation. The pointer device is on the right hand side of the seat.

that the axis of rotation coincided with each participant's naso-occipital axis. This limited the range of possible physical tilt angles to approximately $\pm 13^\circ$. The platform was controlled using Simulink software (The MathWorks, Inc., Natick, Massachusetts, United States). Participants were seated in an automotive style bucket seat (RECARO GmbH, Stuttgart, Germany) that was mounted on top of the platform, and secured with a 5-point safety harness (SCHROTH Safety Products GmbH, Arnsberg, Germany). To minimize head movements, participants also wore a Philadelphia type cervical collar. To ensure that participants could not simply see any tilt of the platform relative to the room, we moved the chair to the edge of the platform. To mask the sounds of the motion platform and the servo-motor, participants wore earplugs with a 33 dB signal-to-noise ratio (Honeywell Safety Products, Roissy, France) as well as a wireless headset (Plantronics, Santa Cruz, California, United States) that actively canceled outside noise, and that played white noise during platform rotations. A photograph of the complete setup is presented in the bottom right panel of Fig. 1.

Task and Stimuli. In each trial, we manipulated the orientation of the visual environment using the alternative-reality system, and the participants' physical orientation by tilting the motion platform. Participants were tasked to align a pointer device with the perceived physical 'up' (the negative of the perceived tilt) on a large number of experimental trials. This method is known as setting the Subjective Haptic Vertical (SHV).

The pointer device consisted of a 15 cm stainless steel rod mounted to a potentiometer. About $\frac{1}{5}$ of the length of the rod extended above its center of rotation. The short hand was to be interpreted as the pointer's top-end, and had to be pointed upwards. The pointer device was free of discontinuities, was not affected by rotation relative to gravity, and provided a $< 0.1^\circ$ resolution. Participants registered a pointer setting as a response by pressing a button at the base of the rod.

There were 25 experimental conditions, comprising all possible combinations of five visual (θ_v) and five physical (θ_l) roll-tilt angles. The specifics of the stimuli were adjusted in three subsequent iterations of the experiment. These adjustments were made to improve discriminability of the statistical models. In the first iteration of the experiment (participants 1–19), θ_v and θ_l both had values of $[-10, -5, 0, 5, 10]^\circ$, where 0° corresponds to the gravitational vertical (maximum discrepancy $\pm 20^\circ$); in the second iteration (participants 20–28), the range of angles was slightly inflated, to $[-13, -6.5, 0, 6.5, 13]^\circ$ (maximum discrepancy $\pm 26^\circ$); and in the final iteration

Iteration	Δ BIC				
	CC _V	CC _I	SS	FF	CI
1	9358.77	478.37	363.75	0.28	0
2	2951.60	278.28	142.63	16.33	0
3	3590.52	2314.98	268.46	426.58	0

Table 1. Overall Δ BIC scores for the three iterations of the experiment. Δ BIC values are calculated as the difference between the BIC score obtained for the model of the corresponding column and the best fitting model. Overall BIC scores were calculated using the sums (over participants) of the model log-likelihoods, the number of parameters, and the number of observations.

(participants 29–36), θ_V had values of $[-50, -25, 0, 25, 50]^\circ$, and θ_I had values of $[-10, -5, 0, 5, 10]^\circ$ (maximum discrepancy $\pm 60^\circ$). In the first and second iteration of the experiment, there were 16 repetitions of each condition, totaling 400 trials; in the third iteration there were 21 repetitions per condition, totaling 525 trials. Note that: θ_V is equal to the sum of the physical tilt and tilt of the camera setup ($\theta_I = \theta_I + \theta_{\text{HMD}}$); equal values for the visual and physical roll tilt angle indicate that the camera was aligned with the participants' physical orientation relative to gravity (i.e., the visual and inertial cue were congruent); and positive angles correspond to clockwise rotations.

To ensure the independence of trials, we tested three methods of transitioning from the visual and physical orientations on one trial to the next during the first iteration of the experiment. For the first nine participants, the camera was turned off while the platform was moved. For the latter five of these participants, heave and sway vibrations were added to the motion profile. These vibrations were in the range of 4–8 Hz and had a root mean square of approximately 0.1 m/s². These vibrations are comparable to road rumble. For the remainder of the participants, the camera was always on during transitions, and there were no vibrations. Regardless of the transition method, the velocity profile of the roll-rotation followed a raised cosine bell, with a duration that was randomly varied between 3–4 s. We did not find any differences in the results of these different subgroups.

Including instructions and 5-minute breaks every 15 minutes, the experiment lasted between two and three hours.

Models of Spatial Orientation. The aim of the present modeling efforts is to assess the tenability of a number of prominent theories on how the Central Nervous System (CNS) constructs perceptions of upright under conditions of uncertainty about the causality of potential cues on the upright.

We postulate that a response R reflecting the perceived upright is equal to the negative of the final tilt estimate r ($R = -r$), and that r is constructed from visual (V) and inertial (I) sensory estimates of orientation (x_V, x_I). The visual system can generate estimates of orientation using polarity information that is present in the optical array (e.g., blue sky/green grass; objects lying on a shelf). Our body's inertial sensors comprise the vestibular system of the inner ear and various kinds of other sensory neurons distributed throughout the body. Because all these neurons are, either directly or indirectly, responsive to accelerations, we treat them as one single inertial system. The inertial sensory system can generate estimates of orientation by identifying the direction of gravity.

For either sensory modality $m \in \{V, I\}$, we assume orientation estimates x_m are realizations of a random variable that is a possibly distorted version of the respective actual tilt θ_m . We further assume that for the presently investigated range of orientations around the true gravitational upright, the noise can be approximated by a Gaussian distribution with standard deviation σ_m , and that the distortion can be expressed with a scaling parameter β_m .

$$P(x_m|\theta_m) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left(-\frac{(x_m - \beta_m\theta_m)^2}{2\sigma_m^2}\right). \quad (1)$$

Note that roll-tilt is a circular variable, which should ideally be represented by circular distributions^{26,29,30}. Predictions on perception from cue combination (fusion) models using circular distributions diverge from those of models using Normal distributions when realizations of variables cross the extremes of the circle ($\pm 180^\circ$), and as a function of intersensory discrepancy. More specifically, whereas sensory weightings and the standard deviation of the integrated percept are unaffected by discrepancies in models based on Normal distributions, discrepancies bias the integrated percept towards the more certain sensory estimate in a model using Von Mises distributions, and the uncertainty of the integrated estimate increases as a function of the size of intersensory discrepancy. For a detailed account on cue combination for circular variables and differences between model predictions see Murray and Morgenstern²⁹. We evaluated differences between predictions of a fusion model using Normal and Von Mises distributions using the average values for the standard deviations of the visual and inertial estimates (as per Table 2) and the maximum discrepancy from the corresponding iteration. The overlap between response densities according to the two different models, expressed as the Bhattacharyya distance³¹, was consistently above 0.99, where 1 is perfect overlap, and 0 is no overlap at all. The predictions on means differed by 0.036°, 0.039°, and 0.498°; and the standard deviations by 0.027°, 0.072°, and 0.820°, for the three iterations, respectively. Because the differences were negligible, we used Gaussian distributions. This allowed us to formulate analytical expressions for the models and to fit them to the data without the need to resort to simulations and numerical integration.

We corrected for constant offset in the responses by calibrating the rod zero-position with the subjective upright at the outset of the experiment, and by subtracting the mean from the data.

Parameter	Model					
	CC _V	CC _I	SS	FF	CI	bounds
Iteration 1						
β_V	0.12 (0.15)		0.57 (0.47)	0.52 (0.66)	0.82 (1.54)	[-5:5]
σ_V	7.60 (4.39)		4.59 (5.08)	8.13 (6.45)	7.65 (8.83)	[1:∞]
β_I		0.86 (0.59)	0.97 (0.57)	1.35 (0.69)	1.25 (0.94)	[-5:5]
σ_I		4.37 (1.82)	3.97 (1.66)	5.13 (2.43)	4.12 (2.32)	[1:∞]
P(V)			0.03 (0.14)		0.50 (0.41)	[0:1]
P(C)					0.82 (0.39)	[0:1]
Iteration 2						
β_V	0.14 (0.18)		0.36 (0.85)	0.43 (0.35)	0.77 (1.92)	[-5:5]
σ_V	8.35 (3.33)		2.99 (3.49)	8.91 (3.59)	10.98 (8.25)	[1:∞]
β_I		0.69 (0.37)	0.80 (0.36)	1.12 (0.49)	1.09 (0.40)	[-5:5]
σ_I		5.84 (2.44)	5.23 (1.40)	7.47 (3.80)	5.43 (1.65)	[1:∞]
P(V)			0.07 (0.24)		0.50 (0.39)	[0:1]
P(C)					0.27 (0.37)	[0:1]
Iteration 3						
β_V	0.22 (0.18)		0.58 (0.13)	0.77 (0.56)	0.66 (0.71)	[-5:5]
σ_V	13.91 (6.43)		8.56 (3.26)	17.63 (13.53)	10.94 (8.37)	[1:∞]
β_I		1.46 (0.89)	1.98 (0.74)	2.17 (0.71)	2.68 (1.17)	[-5:5]
σ_I		15.36 (5.68)	8.04 (2.78)	14.89 (5.33)	12.87 (7.72)	[1:∞]
P(V)			0.26 (0.32)		0.21 (0.36)	[0:1]
P(C)					0.05 (0.32)	[0:1]

Table 2. Median parameter values over participants (and standard deviation). Values are split per experiment iteration. The lower bounds of 1 for σ_V , σ_I were chosen to prevent cases where fitting of the mixture models would result in explanation of a single outlier with a dedicated component with near-zero standard deviation.

We do not have access to the sensory estimates x_m , and are interested in the probability of the tilt estimate r given stimuli θ_V, θ_I . This probability is

$$\begin{aligned} P(r|\theta_V, \theta_I) &= \iint_{-\infty}^{\infty} P(r, x_V, x_I|\theta_V, \theta_I) dx_V dx_I \\ &= \iint_{-\infty}^{\infty} P(r|x_V, x_I, \theta_V, \theta_I) P(x_V, x_I|\theta_V, \theta_I) dx_V dx_I. \end{aligned} \quad (2)$$

We assume that the sensory estimates for different modalities are generated independently. Therefore, $P(x_V, x_I|\theta_V, \theta_I) = P(x_V|\theta_V)P(x_I|\theta_I)$ and the equation above can be rewritten as

$$P(r|\theta_V, \theta_I) = \iint_{-\infty}^{\infty} P(r|x_V, x_I, \theta_V, \theta_I) P(x_V|\theta_V) P(x_I|\theta_I) dx_V dx_I. \quad (3)$$

Since x_V and x_I is the only information available to the observer, r will be conditionally independent of other variables apart from x_V and x_I , i.e.,

$$P(r|x_V, x_I, \theta_V, \theta_I) = P(r|x_V, x_I). \quad (4)$$

By taking this into account and substituting (1) into (3), we get

$$\begin{aligned} P(r|\theta_V, \theta_I) &= \frac{1}{2\pi\sigma_V\sigma_I} \iint_{-\infty}^{\infty} P(r|x_V, x_I) \\ &\quad \times \exp\left(-\frac{(x_I - \beta_I\theta_I)^2}{2\sigma_I^2} - \frac{(x_V - \beta_V\theta_V)^2}{2\sigma_V^2}\right) dx_V dx_I. \end{aligned} \quad (5)$$

Because a participant is unaware of θ_m and β_m , the final tilt estimate generation model $P(r|x_m)$ uses a different sensory estimate (x_m) generation model than the true sensory estimate generation model.

From a participant's perspective, the likelihood of the sensory estimate given any orientation Θ being the true orientation can be expressed as

$$P(x_m|\Theta) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left(-\frac{(x_m - \Theta)^2}{2\sigma_m^2}\right). \quad (6)$$

Moreover, we consider the notion that the CNS includes a-priori beliefs about Θ , namely that we are usually upright, in the construction of this percept. We choose the long body axis as the reference (0°) for other angles. Consequently, we define a prior belief of the following form:

$$P(\Theta) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{\Theta^2}{2\sigma_0^2}\right), \quad (7)$$

where σ_0 is the distribution's standard deviation, which represents the strength of the prior belief. In accordance with the literature, we refer to this prior as the 'idiotropic prior'.

Various general strategies have been proposed on how the CNS may construct final tilt estimates from the multisensory estimates and prior beliefs. Below, we provide mathematical formulations of prominent strategies.

Cue Capture. According to Cue Capture (CC) models, perception of specific environmental properties is dominated by a single sensory modality³²⁻³⁴. In the present case, there are two such possibilities: perception of the upright is dominated by either visual or inertial information. A prior belief that the upright aligns with the long body axis interacts with the sensory information according to Bayes' rule. The posterior probability of Θ given either sensory estimate x_m is then given by:

$$\begin{aligned} P(\Theta|x_m) &= \frac{P(x_m|\Theta)P(\Theta)}{P(x_m)} \propto \exp\left(-\frac{(x_m - \Theta)^2}{2\sigma_m^2}\right) \exp\left(-\frac{\Theta^2}{2\sigma_0^2}\right) \\ &= \exp\left(-\frac{(x_m - \Theta)^2}{2\sigma_m^2} - \frac{\Theta^2}{2\sigma_0^2}\right). \end{aligned} \quad (8)$$

Consistent with the literature, we assume that for individual trials r is the mode of this posterior distribution (i.e., the Maximum-A-Posteriori estimate, MAP)

$$r_{CC}(x_m) = \operatorname{argmax}_{\Theta} P(\Theta|x_m) = \frac{\sigma_0^2 x_m}{\sigma_0^2 + \sigma_m^2} = \alpha_m x_m \quad (9)$$

where $\alpha_m = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_m^2}$. The corresponding PDF can be expressed as

$$P(r_{CC}|x_m) = \delta(r_{CC} - \alpha_m x_m) \quad (10)$$

in which $\delta(\cdot)$ is Dirac's delta function. By substituting (10) into (5) we obtain

$$\begin{aligned} P(r_{CC}|\theta_m) &= \frac{1}{\sqrt{2\pi}\sigma_m} \int_{-\infty}^{\infty} \delta(r_{CC} - \alpha_m x_m) \exp\left(-\frac{(x_m - \beta_m \theta_m)^2}{2\sigma_m^2}\right) dx_m \\ &= \frac{1}{\sqrt{2\pi}\alpha_m\sigma_m} \exp\left(-\frac{(r_{CC} - \alpha_m\beta_m\theta_m)^2}{2\alpha_m^2\sigma_m^2}\right). \end{aligned} \quad (11)$$

Switching Strategy. The CC models can be considered special cases of the Switching Strategy (SS) model^{21,34}. The SS model essentially combines the two alternative CC models: the CNS constructs r for each sensory modality as in the Cue Capture models, but randomly chooses either modality as dominant source on a trial-by-trial basis:

$$r_{SS}(x_V, x_I) = \begin{cases} \alpha_V x_V, & \text{with probability } P(V) \\ \alpha_I x_I, & \text{with probability } 1 - P(V) \end{cases} \quad (12)$$

with α_m as before, with the corresponding probability density function

$$P(r_{SS}|x_I, x_V) = P(V)\delta(r_{SS} - \alpha_V x_V) + (1 - P(V))\delta(r_{SS} - \alpha_I x_I). \quad (13)$$

Filling in (13) into (5), we obtain the likelihood of the responses given the stimuli $P(r_{SS}|\theta_V, \theta_I)$:

$$\begin{aligned} P(r_{SS}|\theta_V, \theta_I) &= \frac{P(V)}{\sqrt{2\pi}\alpha_V\sigma_V} \exp\left(-\frac{(r_{SS} - \alpha_V\beta_V\theta_V)^2}{2\alpha_V^2\sigma_V^2}\right) \\ &\quad + \frac{1 - P(V)}{\sqrt{2\pi}\alpha_I\sigma_I} \exp\left(-\frac{(r_{SS} - \alpha_I\beta_I\theta_I)^2}{2\alpha_I^2\sigma_I^2}\right). \end{aligned} \quad (14)$$

Forced Fusion. In the present formulation of the Forced Fusion (FF) model, it is assumed that visual and inertial sensory estimates are independent from each other, and that both are always interpreted as cues to orientation⁹⁻¹³. The posterior probability of Θ can be expressed as:

$$\begin{aligned}
 P(\Theta|x_V, x_I) &= \frac{P(x_V, x_I|\Theta)P(\Theta)}{P(x_V, x_I)} \\
 &\propto \exp\left(-\frac{(x_V - \Theta)^2}{2\sigma_V^2}\right)\exp\left(-\frac{(x_I - \Theta)^2}{2\sigma_I^2}\right)\exp\left(-\frac{\Theta^2}{2\sigma_0^2}\right) \\
 &= \exp\left(-\frac{(x_V - \Theta)^2}{2\sigma_V^2} - \frac{(x_I - \Theta)^2}{2\sigma_I^2} - \frac{\Theta^2}{2\sigma_0^2}\right).
 \end{aligned}
 \tag{15}$$

The final tilt estimate r_{FF} is again the mode of the posterior distribution (the MAP):

$$r_{FF}(x_V, x_I) = \underset{\Theta}{\operatorname{argmax}} \operatorname{Pr}(\Theta|x_V, x_I) = \alpha_{0V}x_V + \alpha_{0I}x_I, \tag{16}$$

with $\alpha_{0V} = \frac{\sigma_0^2\sigma_I^2}{\sigma_0^2\sigma_I^2 + \sigma_0^2\sigma_V^2 + \sigma_I^2\sigma_V^2}$, and $\alpha_{0I} = \frac{\sigma_0^2\sigma_V^2}{\sigma_0^2\sigma_I^2 + \sigma_0^2\sigma_V^2 + \sigma_I^2\sigma_V^2}$. Similar to (10), the posterior PDF can be expressed as

$$P(r_{FF}|x_V, x_I) = \delta(r_{FF} - \alpha_{0V}x_V - \alpha_{0I}x_I). \tag{17}$$

By substituting (17) into (5) and subsequently simplifying, we obtain

$$\begin{aligned}
 P(r_{FF}|\theta_V, \theta_I) &= \frac{1}{2\pi\sigma_V\sigma_I} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(r_{FF} - \alpha_{0V}x_V - \alpha_{0I}x_I) \\
 &\quad \exp\left(-\frac{(x_V - \beta_V\theta_V)^2}{2\sigma_V^2} - \frac{(x_I - \beta_I\theta_I)^2}{2\sigma_I^2}\right) dx_V dx_I \\
 &= \frac{1}{\sqrt{2\pi}\sigma_{FF}} \exp\left(-\frac{(r_{FF} - \mu_{FF})^2}{2\sigma_{FF}^2}\right),
 \end{aligned}
 \tag{18}$$

where

$$\mu_{FF} = \beta_I\theta_I\alpha_{0I} + \beta_V\theta_V\alpha_{0V}, \tag{19}$$

$$\sigma_{FF}^2 = \sigma_I^2\alpha_{0I}^2 + \sigma_V^2\alpha_{0V}^2. \tag{20}$$

Causal Inference Model. In the models presented above, it is assumed that the CNS either segregates (CC, SS) or fuses multisensory information (FF). CI models allow segregation and fusion to occur in tandem; the estimates generated by the different strategies are treated as intermediate estimates, and a final estimate is constructed by taking into account the probability that the internal estimates share a common cause (C), favoring FF; or have independent causes (\bar{C}), favoring SS^{22,23,25,26,30}.

The probability of a common cause given the sensory estimates is

$$\begin{aligned}
 P(C|x_V, x_I) &= \frac{P(x_V, x_I|C)P(C)}{P(x_V, x_I)} \\
 &= \frac{P(x_V, x_I|C)P(C)}{P(x_V, x_I|C)P(C) + P(x_V, x_I|\bar{C})(1 - P(C))}
 \end{aligned}
 \tag{21}$$

where $P(C)$ is a free parameter that represents a prior tolerance for discrepancies. The likelihood of the sensory estimates x_V, x_I given a common cause C is

$$P(x_V, x_I|C) = \int_{-\infty}^{\infty} P(x_V, x_I|\Theta)P(\Theta)\Theta. \tag{22}$$

where $P(\Theta)$ is the idiotropic prior. $P(x_V, x_I|\Theta)$ is the likelihood of x_V, x_I given some common orientation Θ . This becomes

$$\begin{aligned}
 P(x_V, x_I|C) &= \frac{1}{(\sqrt{2\pi})^3\sigma_V\sigma_I\sigma_0} \int_{-\infty}^{\infty} \exp\left(-\frac{(x_V - \Theta)^2}{2\sigma_V^2} - \frac{(x_I - \Theta)^2}{2\sigma_I^2} - \frac{\Theta^2}{2\sigma_0^2}\right) d\Theta \\
 &= \frac{1}{2\pi\sqrt{\sigma_V^2\sigma_I^2 + \sigma_V^2\sigma_0^2 + \sigma_I^2\sigma_0^2}} \exp\left(-\frac{1}{2} \frac{(x_V - x_I)^2\sigma_0^2 + x_V^2\sigma_I^2 + x_I^2\sigma_V^2}{\sigma_V^2\sigma_I^2 + \sigma_V^2\sigma_0^2 + \sigma_I^2\sigma_0^2}\right).
 \end{aligned}
 \tag{23}$$

For an interpretation of independent causes, r will be based on either the visual or the inertial estimate. We assume that the same idiotropic prior interacts with both sensory estimates, as only one of them will ultimately be treated as informative of body tilt.

$$\begin{aligned}
P(x_V, x_I | \bar{C}) &= \frac{1}{2\pi\sigma_V\sigma_0} \int_{-\infty}^{\infty} \exp\left[-\frac{(x_V - \Theta)^2}{2\sigma_V^2} - \frac{\Theta^2}{2\sigma_0^2}\right] d\Theta \\
&\quad \times \frac{1}{2\pi\sigma_I\sigma_0} \int_{-\infty}^{\infty} \exp\left[-\frac{(x_I - \Theta)^2}{2\sigma_I^2} - \frac{\Theta^2}{2\sigma_0^2}\right] d\Theta \\
&= \frac{1}{2\pi\sqrt{(\sigma_V^2 + \sigma_0^2)(\sigma_I^2 + \sigma_0^2)}} \exp\left[-\frac{1}{2}\left(\frac{x_V^2}{\sigma_V^2 + \sigma_0^2} + \frac{x_I^2}{\sigma_I^2 + \sigma_0^2}\right)\right].
\end{aligned} \tag{24}$$

As in^{22,24}, r_{CI} is a weighted average of the intermediate tilt estimates according to the SS strategy r_{SS} and according to the FF strategy r_{FF} , with weights proportional to the respective probability of the causal structures

$$r_{CI}(x_V, x_I) = P(\bar{C}|x_V, x_I)r_{SS}(x_V, x_I) + P(C|x_V, x_I)r_{FF}(x_V, x_I) \tag{25}$$

where $P(\bar{C}|x_V, x_I) = 1 - P(C|x_V, x_I)$. r_{FF} is a deterministic function of (x_V, x_I) and r_{SS} is a random variable which can take one of the two values according to (12). Therefore, r_{CI} can be written as

$$\begin{aligned}
r_{CI}(x_V, x_I) &= P(C|x_V, x_I)r_{FF}(x_V, x_I) \\
&\quad + (1 - P(C|x_V, x_I)) \begin{cases} \alpha_V x_V, & \text{with probability } P(V) \\ \alpha_I x_I, & \text{with probability } 1 - P(V) \end{cases}
\end{aligned} \tag{26}$$

which can be expressed as the density function

$$P(r_{CI}|x_V, x_I) = P(V)\delta(r_{CI} - A(x_V, x_I)) + (1 - P(V))\delta(r_{CI} - B(x_V, x_I)) \tag{27}$$

where

$$\begin{aligned}
A(x_V, x_I) &= P(C|x_V, x_I)r_{FF}(x_V, x_I) + (1 - P(C|x_V, x_I))\alpha_V x_V \\
B(x_V, x_I) &= P(C|x_V, x_I)r_{FF}(x_V, x_I) + (1 - P(C|x_V, x_I))\alpha_I x_I.
\end{aligned}$$

By substituting the expression for r_{FF} (16) into the equations above and doing some transformations we obtain

$$\begin{aligned}
A(x_V, x_I) &= P(C|x_V, x_I)(\alpha_{0V}x_V + \alpha_{0I}x_I - \alpha_V x_V) + \alpha_V x_V \\
B(x_V, x_I) &= P(C|x_V, x_I)(\alpha_{0V}x_V + \alpha_{0I}x_I - \alpha_I x_I) + \alpha_I x_I.
\end{aligned}$$

By replacing the final tilt estimate generating model $P(r|x_V, x_I)$ in (5) with (27), we can obtain the likelihood function for r given the stimuli. However, due to the $P(C|x_V, x_I)$ expression, the integral in (5) cannot be represented in a closed form. To resolve this issue, we linearize $A(x_V, x_I)$ and $B(x_V, x_I)$ at $x_V = x_{V0} = \beta_V \theta_V$, $x_I = x_{I0} = \beta_I \theta_I$. For A, B , we obtain

$$A(x_V, x_I) \approx \frac{A(x_{V0}, x_{I0})}{A_0(\theta_V, \theta_I)} + \frac{dA}{dx_V}(x_{V0}, x_{I0})(x_V - x_{V0}) + \frac{dA}{dx_I}(x_{V0}, x_{I0})(x_I - x_{I0}) \tag{28}$$

$$B(x_V, x_I) \approx \frac{B(x_{V0}, x_{I0})}{B_0(\theta_V, \theta_I)} + \frac{dB}{dx_V}(x_{V0}, x_{I0})(x_V - x_{V0}) + \frac{dB}{dx_I}(x_{V0}, x_{I0})(x_I - x_{I0}). \tag{29}$$

If we use the approximations for $A(x_V, x_I)$, $B(x_V, x_I)$ in (27), we can solve the integrals in (5). This yields

$$P(r_{CI}|\theta_V, \theta_I) = P(V) \left(\frac{1}{\sqrt{2\pi v_V}} \right) \exp\left[-\frac{(r - A_0(\theta_V, \theta_I))^2}{2\sigma_A^2}\right] \tag{30}$$

$$+ (1 - P(V)) \left(\frac{1}{\sqrt{2\pi v_I}} \right) \exp\left[-\frac{(r - B_0(\theta_V, \theta_I))^2}{2\sigma_B^2}\right] \tag{31}$$

with $\sigma_A^2 = a_V^2\sigma_V^2 + a_I^2\sigma_I^2$ and $\sigma_B^2 = b_V^2\sigma_V^2 + b_I^2\sigma_I^2$.

We validated the linear approximation by comparing the results with those of performing numerical integrations for the initial few participants.

Data analysis. The parameters to account for distortions in perception (β_V, β_I); the standard deviations of the sensory estimates (σ_V, σ_I); and the mixture and ‘tolerance for discrepancies’ prior parameters ($P(V), P(C)$) were treated as free parameters, resulting in a total of two to six free parameters, depending on the model. The standard deviation of the idiosyncratic prior σ_0 was initially also included as a free parameter. However, doing so was found to result in problems with optimization convergence and generally yielded extremely large values for this parameter,

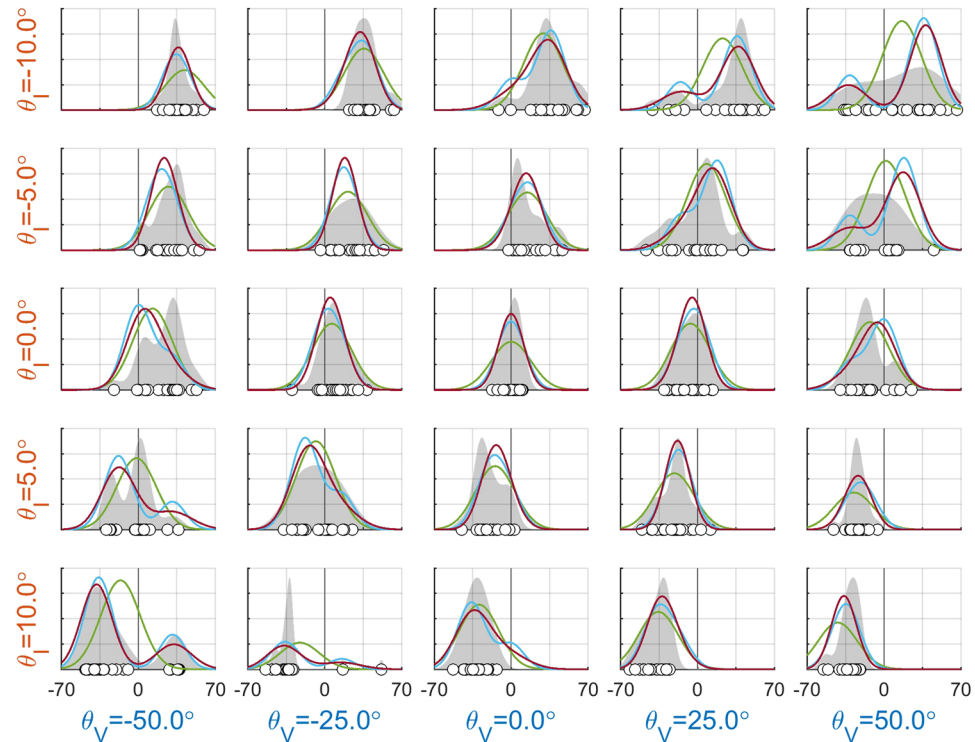


Figure 2. Overview of the results for an example participant (31). Each panel shows the data of a particular experimental condition. Responses (white dots) reflect the negative of the perceived tilt. The gray-shaded areas show the corresponding kernel density estimates. The thin black line at 0° is the Earth-vertical. The colored lines represent the response densities according to the SS (blue), FF (green), and CI (red) models that allowed for distortion in perceptions. Note how the CI model allows for behaviors in between the FF and SS models.

suggesting that its effect on the perception of verticality was negligible. The existence of an idiotropic vector was proposed to account for biases in the SVV⁸, but the findings of previous studies found this prior not to affect the SHV^{35,36}. Consequently, we fixed the value of σ_0 to 100. This renders the prior effectively uninformative, but we chose to include it for consistency with the literature.

We fitted the CC and FF models by minimizing the negative log-likelihood of the tilt estimates ($r = -R$) given the model $-\sum_{i=1}^n \log(\Pr(r_{\text{model}_i} | \theta_V, \theta_I))$, using the `fmincon` routine in MATLAB. The `fmincon` routine was not suitable to fit the SS and CI models: these are mixture models, and directly maximizing the likelihood can lead to numerical issues. We therefore applied the Expectation-Maximization (EM) algorithm to fit these models³⁷. In the EM-algorithm, membership of mixture components is treated as a latent variable. The model likelihood is maximized iteratively, by repeating a set of two steps: first, the probability of each observation belonging to either component of the mixture is determined given an initial set of parameters. Second, the model parameters are re-estimated, while taking the probability of component membership calculated in the previous step into account. To re-estimate the parameters in the second step, we again minimized the model negative log-likelihoods using the MATLAB `fmincon` routine. The iterations were terminated when the change in model likelihood was smaller than 1×10^{-6} .

To determine which model best approximated participant responses, we compared model Bayesian Information Criterion (BIC) scores³⁸. The BIC is a penalized likelihood score, taking into account the number of observations and the number of free parameters in each model. The model with the lowest BIC score is considered the best in an absolute sense. Differences in model BIC scores (ΔBIC) between 0–2; 2–6; 6–10 are considered negligible, positive, and strong evidence, respectively, and $\Delta\text{BIC} > 10$ are considered decisive evidence.

For each of the models, we evaluated the fit of an additional version, where the values of the β_V and β_I parameters were fixed at a value of 1, reflecting an assumption that perception itself is veridical. Here, distortion of the final response R was attributed to an over- or underestimation of the angle of the rod. This was implemented as a linear transformation of the response random variable: $R = \beta_r r_{\text{model}}$, with $\text{Var}(R) = \beta_r^2 \text{Var}(r_{\text{model}})$. In this version, the scaling parameter (β_r) affects the noise parameters (σ_m). To illustrate, when the response reflects a consistent underestimation of the tilt angle, this will result in $\beta_r < 1$. This also implies that the noise parameter for perceived tilt r must be increased to fit the variability in responses R , compared to the first version of the models. Ultimately, this version of the models was found to provide a reasonable alternative only for the FF model, and we therefore chose not to further consider the findings of this version of the models. The model fits are available as supplementary material, in Tables S8–S14.

It should also be noted that while it is theoretically possible that distortions are introduced both at the perceptual and response levels, it was not possible to estimate both effects simultaneously because distortion parameters at both levels allow similar behavior, resulting in an infinite number of equivalent solutions.

Results

In the following, we separately present the findings on model fits and parameter estimates. As an illustration of the findings, Fig. 2 shows data and model fits for an example participant. Figures showing the data of all individual participants are provided as supplementary material Figs S1–S36.

Model comparisons. In the first iteration of the experiment, where stimuli with discrepancies up to $\pm 20^\circ$ were presented, the evidence did not allow to decide upon a best fitting model, as the results were tied between the FF model (BIC = 44710.67) and the CI model (BIC = 44710.39), resulting in an overall Δ BIC of 0.28, which is considered negligible evidence. Inspection of individual results also revealed a considerable variability between participants with respect to the preferred model: the CC_1 model provided the best fit in six cases; the SS model in one case; the FF model in seven; and the CI model in the five remaining cases.

In the second iteration, where stimuli with discrepancies up to $\pm 26^\circ$ were presented, the evidence favored the CI model, as indicated by a Δ BIC score of 16.33. The individual results however again exhibited variability between participants. Here, the SS model was preferred in one case, the FF model in three, and the CI model in five cases.

The results of the third iteration of the experiment, with discrepancies up to $\pm 60^\circ$, favored the CI model. This was evidenced by a Δ BIC value of 268.46, compared to the runner-up model SS. Individual results were also consistent, providing support for the CI model in seven out of eight cases, and providing support for the SS model in the remaining case (Δ BIC = 4.42, compared to CI). Overall Δ BIC scores for the three iterations are presented in Table 1. The negative log-likelihood and BIC scores obtained for individual participants are presented in supplementary material Tables S1 and S2.

Parameter estimates. Summaries of the obtained estimates for the scaling parameters β_v, β_p , the standard deviations of the sensory estimates σ_v and σ_p , and the mixture and prior parameters P(V) and P(C) for each of the three iterations of the experiment are presented in Table 2. The obtained parameter estimates for each model/individual are available as supplementary material Tables S3–S7.

The median value of scaling parameter β_v varied considerably between models and iterations, ranging from 0.12 (CC_v , iteration 1) to 0.82 (CI, iteration 1). For the CC_v model, the parameter's value can be interpreted as a regression coefficient, and as such indicates a minor effect of vision on tilt estimates. For the latter models, the parameter's overall median value was 0.57. A value below one indicates that the visual tilt was underestimated. The median value for the visual noise parameter σ_v ranged from 2.99 (SS, iteration 2) to 17.63 (FF, iteration 3), and was largest in the third iteration.

The median value for scaling parameter β_p showed a similar variability, ranging from 0.69 (CC_p , iteration 2) to 2.68 (CI, iteration 3), but was generally larger than 1 (overall median value 1.33). For the CC_p model, the parameter can again be interpreted as a regression coefficient, and indicates a larger effect for physical tilt than for visual tilt. The finding that the parameter's value was generally larger than one for the SS, FF, and CI models indicates that physical tilt was overestimated. The median value for the inertial noise parameter σ_p ranged from 3.97 to 15.36, and was also the largest in the third iteration.

Assuming forced fusion, the observed values of the σ_v, σ_p parameters can be translated into relative contributions of visual and inertial information to an integrated estimate. For the FF model, the observed values translate into visual:inertial weights of 0.28:0.72, 0.41:0.59, and 0.42:0.58, for the three iterations, respectively; for the CI model, these weights would correspond to 0.23:0.77, 0.20:0.80, and 0.59:0.41.

Parameter P(V) was generally close to zero for the SS model (medians 0.03, 0.07, 0.26), suggesting that participants generally relied on inertial information (as in the CC_1 model), but occasionally lapsed by relying on visual information. In the CI model, the value of this parameter was larger in the first two iterations (medians 0.50, 0.50), indicating that responses were more likely to reflect visual information when a discrepancy was likely, but closer to the SS model estimate in the final iteration (0.21).

The a-priori belief that signals will have a common cause, reflected by parameter P(C) of the CI model had median values of 0.82, 0.27, 0.05, suggesting that the range of discrepancies affects participants' a-priori tendency to assume a common cause.

Discussion

We investigated how perceptions of upright are constructed under conditions of uncertainty about the veracity of visual and inertial cues. We manipulated the maximum discrepancy between the orientation suggested by these cues in three experimental iterations, with maximum discrepancies increasing from $\pm 20^\circ$ in the first iteration, to $\pm 26^\circ$ in the second, and $\pm 60^\circ$ in the final iteration.

Perceptual bias. The perception of verticality has been shown to be subject to a number of biases. Most notably, distortions are caused by ocular counterrolling (OCR), and hysteresis.

OCR is roll rotation of the eyes in the direction opposite to the inducing stimulus. OCR can be induced by physical as well as visual tilt stimuli (e.g.^{39–42}), and can amount to up to approximately 10% of the roll-stimulus angle^{2,42}. It has been shown that the perceptual system does not correct for such torsional motion⁴³. Our setup did not allow us to measure OCR directly. Instead, we addressed the possibility of systematic distortions by including scaling parameters for the unsensory visual and inertial tilt estimates in the modeling. The visual scaling parameter was found to be generally below 1, which is consistent with the findings on OCR discussed above, although the contribution

of the visual information was scaled down more than the 10% expected from the literature. In contrast, the inertial scaling parameter was generally considerably larger than 1, indicating physical tilt was overestimated. This finding appears to be consistent with the E-effect, which is an apparent overestimation of head tilt for (relatively) small physical tilt angles^{2,44,45}. It is nevertheless surprising to note that the observed overestimations, and subsequently the scaling parameters for physical tilt (β_p), were considerably larger in the third iteration of the experiment than in the first iteration of the experiment, whereas the presented physical tilt angles were equal. Because the only difference in the paradigm between these iterations was an increase in the range of visual tilt angles, we speculate that the range of visually perceived angles may affect expectations on the range of physical tilt angles, and consequently affect the scaling of inertial sensory estimates. This could be interpreted as a cross-modal range effect (see e.g.⁴⁶).

Hysteresis is the phenomenon where the state of a system depends on its previous state. It has been shown that the SHV task is subject to this effect^{36,47}. We tested three different methods of transitioning between stimuli in the first iteration of the experiment to address this effect. For the first method, the camera was turned off during transitions; for the second method, heave and sway vibrations were added to the transitional rotations; and for the third method, the camera was always kept on whereas the vibrations were omitted. The first allows hysteresis effects of the inertial cue, as this cue is always present and could be tracked during transitions. Due to the vibrations, this is not feasible for the second method. Both visual and inertial hysteresis effects are possible for the third method, but the senses present conflicting information because stimuli were presented in random order, with the added requirement that there was always transitional rotation. Ultimately, the manipulations did not appear to affect the results, suggesting that participants were able to evaluate the stimuli independently. We chose the latter method for the remainder of the experiment because the sensory modalities are treated in the most similar way.

Model comparisons. The results of the first and second iterations did not allow us to discriminate between the models with certainty, as the overall evidence was tied between the FF and CI models. On an individual level, the FF and CI models provided the best explanation of the data for an equal number of participants. Preference of the FF model over the CI model would imply that the SHV is constructed by mandatory fusion of multisensory information on spatial orientation. The standard deviations of visual and inertial estimates further indicated that the contribution of the visual cue to the perceived upright was smaller than the contribution of the inertial cue, with approximate relative weights of 0.35:0.65 (visual:inertial). The observed relative weightings resemble those reported in studies on the SVV^{14,18} and SHV⁴⁵, and the consistency of responses with predictions from FF has also been reported by studies assessing the perceptual upright with other methods, such as the Oriented Character Recognition Test (OCHART)^{14,18}; the SVV³⁶; and SBT¹⁷.

Despite these parallels, a conclusion that perceptions of upright are constructed according to FF would be at odds with recent findings on multisensory interactions regarding perception of other environmental properties, such as audio-visual localization tasks (e.g.^{22–25}) and visual-inertial heading estimation^{26,27}, where it was found that the CNS includes assessments of signal causality in the formation of perceptions. Evaluation of the findings of the first two iterations indicated that predictions made by the different models were quite similar, making it difficult to discriminate between them even for the largest discrepancies. To address this, we used the CI parameter estimates obtained in the first and second iteration of the experiment to simulate response distributions, and to determine whether these would diverge more clearly, for even larger discrepancies. Based on the findings of these simulations, we increased the maximum discrepancy to $\pm 60^\circ$, and performed a third iteration of the experiment with eight additional participants. This had the desired effect, as here the models produced distinct predictions. Evaluation of the responses further provided decisive evidence in favor of the CI model. Because CI behaves as FF when discrepancies are small (i.e., when the size of the discrepancies does not clearly exceed the respective sensory noises), the additional findings do not conflict with the cases where the FF model was preferred in the first two iterations of the experiment; FF was the preferred model because it has fewer parameters. We conclude that, consistent with the hypothesis, the CNS incorporates assessments of signal causality in the perception of verticality, but that this effect becomes unambiguous only for large discrepancies.

References

- Howard, I. P., Bergström, S. S. & Ohmi, M. Shape from shading in different frames of reference. *Percept.* **19**, 523–530 (1990).
- Howard, I. P. *Human visual orientation* (John Wiley & Sons, 1982).
- Lackner, J. R. & Graybiel, A. Postural illusions experienced during z-axis recumbent rotation and their dependence upon somatosensory stimulation of the body surface. *Aviat. space, environmental medicine* (1978).
- Lackner, J. R. & Graybiel, A. Some influences of touch and pressure cues on human spatial orientation. *Aviat. space, environmental medicine* (1978).
- Vaitl, D., Mittelstaedt, H. & Baisch, F. Shifts in blood volume alter the perception of posture. *Int. J. Psychophysiol.* **27**, 99–105 (1997).
- Vaitl, D., Mittelstaedt, H., Saborowski, R., Stark, R. & Baisch, F. Shifts in blood volume alter the perception of posture: further evidence for somatic graviception. *Int. J. Psychophysiol.* **44**, 1–11 (2002).
- Mittelstaedt, H. Somatic graviception. *Biol. psychology* **42**, 53–74 (1996).
- Mittelstaedt, H. A new solution to the problem of the subjective vertical. *Naturwissenschaften* **70**, 272–281 (1983).
- Clark, J. J. & Yuille, A. L. *Data fusion for sensory information processing systems*, vol. 105 (Springer Science & Business Media, 1990).
- Landy, M. S., Maloney, L. T., Johnston, E. B. & Young, M. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vis. research* **35**, 389–412 (1995).
- Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
- Hillis, J. M., Ernst, M. O., Banks, M. S. & Landy, M. S. Combining sensory information: mandatory fusion within, but not between, senses. *Sci.* **298**, 1627–1630 (2002).
- Ernst, M. O. & Bühlhoff, H. H. Merging the senses into a robust percept. *Trends in cognitive sciences* **8**, 162–169 (2004).
- Dyde, R. T., Jenkin, M. R. & Harris, L. R. The subjective visual vertical and the perceptual upright. *Exp. Brain Res.* **173**, 612–622 (2006).
- MacNeilage, P. R., Banks, M. S., Berger, D. R. & Bühlhoff, H. H. A Bayesian model of the disambiguation of gravito-inertial force by visual cues. *Exp. Brain Res.* **179**, 263–290 (2007).

16. Vingerhoets, R. A. A., De Vrijer, M., Van Gisbergen, J. A. & Medendorp, W. P. Fusion of visual and vestibular tilt cues in the perception of visual vertical. *J. neurophysiol.* **101**, 1321–1333 (2009).
17. Clemens, I. A., De Vrijer, M., Selen, L. P., Van Gisbergen, J. A. & Medendorp, W. P. Multisensory processing in spatial orientation: an inverse probabilistic approach. *J. Neurosci.* **31**, 5365–5377 (2011).
18. Harris, L. R., Jenkin, M., Jenkin, H., Zacher, J. E. & Dyde, R. T. The effect of long-term exposure to microgravity on the perception of upright. *npj Microgravity* **3**, 3 (2017).
19. Dyde, R. T., Jenkin, M. R., Jenkin, H. L., Zacher, J. E. & Harris, L. R. The effect of altered gravity states on the perception of orientation. *Exp. brain research* **194**, 647–660 (2009).
20. Kaptein, R. G. & Van Gisbergen, J. A. Interpretation of a discontinuity in the sense of verticality at large body tilt. *J. Neurophysiol.* **91**, 2205–2214 (2004).
21. de Winkel, K. N., Clément, G., Groen, E. L. & Werkhoven, P. J. The perception of verticality in lunar and martian gravity conditions. *Neurosci. letters* **529**, 7–11 (2012).
22. Körding, K. P. *et al.* Causal inference in multisensory perception. *PLoS One* **2**, e943 (2007).
23. Sato, Y., Toyozumi, T. & Aihara, K. Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural computation* **19**, 3335–3355 (2007).
24. Beierholm, U., Shams, L., Ma, W. J. & Koerding, K. Comparing bayesian models for multisensory cue combination without mandatory integration. In *Advances in neural information processing systems*, 81–88 (2008).
25. Wozny, D. R., Beierholm, U. R. & Shams, L. Probability matching as a computational strategy used in perception. *PLoS Comput. Biol.* **6**, e1000871 (2010).
26. De Winkel, K. N., Katliar, M. & Bühlhoff, H. H. Causal inference in multisensory heading estimation. *PLoS one* **12**, e0169676 (2017).
27. Acerbi, L., Dokka, K., Angelaki, D. E. & Ma, W. J. Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. *bioRxiv* 150052 (2017).
28. Harris, L. R., Jenkin, M., Dyde, R. T. & Jenkin, H. Enhancing visual cues to orientation: Suggestions for space travelers and the elderly. In *Progress in brain research*, vol. 191, 133–142 (Elsevier, 2011).
29. Murray, R. F. & Morgenstern, Y. Cue combination on the circle and the sphere. *J. vision* **10**, 15–15 (2010).
30. De Winkel, K. N., Katliar, M. & Bühlhoff, H. H. Forced fusion in multisensory heading estimation. *PLoS One* **10**, e0127104 (2015).
31. Bhattacharyya, A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.* **35**, 99–109 (1943).
32. Rock, I. & Victor, J. Vision and touch: An experimentally created conflict between the two senses. *Sci.* **143**, 594–596 (1964).
33. de Winkel, K. N., Weesie, J., Werkhoven, P. J. & Groen, E. L. Integration of visual and inertial cues in perceived heading of self-motion. *J. vision* **10**, 1–1 (2010).
34. De Winkel, K. *et al.* Integration of visual and inertial cues in the perception of angular self-motion. *Exp. brain research* **231**, 209–218 (2013).
35. Bortolami, S. B., Pierobon, A., DiZio, P. & Lackner, J. R. Localization of the subjective vertical during roll, pitch, and recumbent yaw body tilt. *Exp. brain research* **173**, 364–373 (2006).
36. Vingerhoets, R. A. A., Medendorp, W. P. & Van Gisbergen, J. A. Body-tilt and visual verticality perception during multiple cycles of roll rotation. *J. neurophysiology* **99**, 2264–2280 (2008).
37. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the em algorithm. *J. royal statistical society. Ser. B (methodological)* 1–38 (1977).
38. Schwarz, G. Estimating the dimension of a model. *The annals statistics* **6**, 461–464 (1978).
39. Miller, E. F. Counterrolling of the human eyes produced by head tilt with respect to gravity. *Acta oto-laryngologica* **54**, 479–501 (1962).
40. Cheung, B. *et al.* Human ocular torsion during parabolic flights: an analysis with scleral search coil. *Exp. brain research* **90**, 180–188 (1992).
41. Kingma, H., Stegeman, P. & Vogels, R. Ocular torsion induced by static and dynamic visual stimulation and static whole body roll. *Eur. Arc. Oto-rhino-laryngology* **254**, S61–S63 (1997).
42. Bockisch, C. J. & Haslwanter, T. Three-dimensional eye position during static roll and pitch in humans. *Vis. research* **41**, 2127–2137 (2001).
43. Wade, S. W. & Curthoys, I. S. The effect of ocular torsional position on perception of the roll-tilt of visual stimuli. *Vis. research* **37**, 1071–1078 (1997).
44. Müller, G. E. *Über das Aubertsche Phänomen* (JA Barth, 1915).
45. Barnett-Cowan, M. & Harris, L. R. Perceived self-orientation in allocentric and egocentric space: effects of visual and physical tilt on saccadic and tactile measures. *Brain research* **1242**, 231–243 (2008).
46. Petzschner, F. H. & Glasauer, S. Iterative bayesian estimation as an explanation for range and regression effects: a study on human path integration. *J. Neuroscience* **31**, 17220–17229 (2011).
47. Schuler, J. R., Bockisch, C. J., Straumann, D. & Tarnutzer, A. A. Precision and accuracy of the subjective haptic vertical in the roll plane. *BMC neuroscience* **11**, 83 (2010).

Author Contributions

The study was conceived and designed by K.N.d.W. Apparatus were developed by K.N.d.W. and D.D. Data acquisition was performed by K.N.d.W. and D.D. The data analysis was developed by K.N.d.W. and M.K. Interpretation of the data and drafting of the manuscript was done by K.N.d.W. All authors contributed to critical revision of the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-23838-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018